UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/944,332	08/30/2001	Liqin Shen	JP920000191US1 (590.079)	1865

35195	7590	08/21/2006

FERENCE & ASSOCIATES
409 BROAD STREET
PITTSBURGH, PA 15143

EXAMINER
HAN, QI

ART UNIT	PAPER NUMBER
2626	

DATE MAILED: 08/21/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

	Application No.	Applicant(s)
Office Action Summary	09/944,332	SHEN ET AL.
	Examiner	Art Unit
	Qi Han	2626

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *10 July 2006*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-19* is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-19* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☒ All b)☐ Some * c)☐ None of:

      1.☒ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☒ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
    Paper No(s)/Mail Date _____.

4)☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____ .

5)☐ Notice of Informal Patent Application (PTO-152)

6)☐ Other: _____.

DETAILED ACTION

1.   The text of those sections of Title 35, U.S. Code not included in this action can be found

in a prior Office action.

Continued Examination Under 37 CFR 1.114

2.   A request for continued examination under 37 CFR 1.114, including the fee set forth in

37 CFR 1.17(e), was filed in this application after final rejection. Since this application is

eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e)

has been timely paid, the finality of the previous Office action has been withdrawn pursuant to

37 CFR 1.114.

Priority

3.   Receipt is acknowledged of papers submitted under 35 U.S.C. 119(a)-(d), which papers

have been placed of record in the file.

Response to Amendment

4.   This communication is responsive to the applicant's amendment and RCE examination

both filed on 07/10/2006. The applicant(s) amended claims 1, 10 and 19 (see the amendment:

pages 2-6).

The examiner withdraws the disclosure objection regarding the term "ANWE", because

the applicant amended the corresponding content in the specification (see page 2 of the

amendment filed on 04/10/2006).

Response to Arguments

5.   Applicant's arguments filed on 07/10/2006 with respect to claims 1-19 have been fully

considered but are moot in view of the new ground(s) of rejection, since the amended

independent claims introduce new issue/new subject matter and change the scope of the claims

(see detail in the claim rejection below). The response to the applicant's arguments is also

directed to the claim rejection below, because the arguments (see the amendment: pages 7-10)

are based on the amended claims.

Specification

6.   The disclosure is objected to because of the following (use the same reference numbers):

a.   On page 6, line 7, regarding the content "length (S)-N is $N(N+1)/2$", even though

the applicant amended it as "**length (S)=N is $N(N+1)/2$**" (see page 2 of the amendment

filed on 09/21/2005) and further suggested that the symbol N is treated as the same as

symbol N (see page 8 of the amendment filed on 04/10/2006), it is still logically and/or

mathematically incorrect. Appropriate correction is required.

Claim Rejections - 35 USC § 112

The following is a quotation of the first paragraph of 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

7.   Claims 1, 10 and 19 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention.

Regarding claim 1, the new amended limitation "wherein the segmenting and the splitting is not dependent upon word boundaries" introduces new subject matter, because the limitation is not specifically described in the original specification.

Regarding claims 10 and 19, the rejection is based on the same reason as described for claim 1, because the claims recite the same or similar limitation as claim 1.

The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

8.   Claims 1, 10 and 19 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Regarding claim 1, the limitation "the segmenting and the splitting is not dependent upon word boundaries" is indefinite, because the applicant introduces contradictory statements. For example, it is unclear whether the claimed invention is **only dealing with** processing Chinese or Japanese languages, for which "there is no word boundary in written languages" (see the

.specification: page 1, line 1), **or dealing with** "any language" as presented in the applicant's

arguments (see the amendment: page 8, paragraph 3), wherein the statements conflict with each

other.

Regarding claims 10 and 19, the rejection is based on the same reason as described for

claim 1, because the claims recite the same or similar limitation as claim 1.

Claim Rejections - 35 USC § 103

9.   Claims 1-3, 6-12 and 15-19 are rejected under 35 U.S.C. 103(a) as being unpatentable

over Wang et al. (US 6,904,402 B1) hereinafter referenced as Wang, in view of Razin et al. (US

6,098,034) hereinafter referenced as Razin and Yang et al. ("statistics-based segment pattern

lexicon—a new direction for Chinese language modeling", 0-7803-4428-6/98, IEEE, pp 169-

172) hereinafter referenced as Yang.

As per **claim 1**, as best understood in view of the rejection under 35 USC 112 1st and 2nd,

(see above), Wang discloses system and iterative method for lexicon, segmentation and language

model joint optimization (title), comprising:

"segmenting a cleaned corpus to form a segmented corpus", (Fig. 5 and column 9, lines

36-44, 'segmentation', 'the received corpus is built', 'pre-processed to remove some obvious

illogical words (so as to provide cleaned corpus)');

"splitting the segmented corpus to form sub strings, and counting the occurrences of each

sub strings appearing in the corpus" (column 1, lines 45-60, 'a textual corpus is dissected

(interpreted as split) into a plurality of items (sub strings)' and 'counts the number of

occurrences of a particular item (word, character, etc.)'); and

Even though Wang further suggests that 'the items of the corpus' having low occurrence

frequency 'may be pruned' (column 7, lines 27-29) and 'counting the occurrence of strings of

characters' (corresponding to new words and is capable of outputting), Wang does not expressly

disclose "**filtering** out false candidates to output new words". However, this feature is well

known in the art as evidenced by Razin who, in the same field of endeavor, discloses method for

standardizing phrasing in a document (title), comprising 'filtering the preliminary list of

extracted phrases (candidates) to create (output) a final list of extracted phrases (corresponding

new words)' (Fig. 2 and column 29, lines 55-56). Therefore, it would have been obvious to one

of ordinary skill in the art at the time the invention was made to modify Wang by specifically

providing filtering a set of extracted phrases and creating (output) final phrases list, as taught by

Razin, for the purpose (motivation) of obtaining extracted words constituting significant user

phrases (or new words) (Razin: column 2, lines 46-47).

It is noted Wang in view Razin does not **expressly** disclose "the segmenting and the

splitting is not dependent upon word boundaries". However, the feature is well known in the art

as evidenced by Yang who, in the same field of endeavor, discloses 'statistics-based segment

pattern lexicon—a new direction for Chinese language modeling' (title), teaches that since 'there

are no "blanks" in Chinese sentences serving as word boundaries, ...the "word" in Chinese are

actually not well defended' (abstract), so that the elements in the lexicon called 'segment pattern

of characters' 'should be extracted form the training corpus (corresponding to clean corpus) from

the training corpus by statistical approach (that is not dependent upon word boundaries)' (page

169, right column, paragraph 3). Therefore, it would have been obvious to one of ordinary skill

in the art at the time the invention was made to modify Wang in view of Razin by specifically

providing a statistical approach of segmentation of Chinese language with segment patterns of

characters for training corpus, as taught by Yang, for the purpose (motivation) of minimizing the

overall perplexity for the segmentation (Yang: abstract).

In addition, in another view of disclosure of Wang and Razin, Wang further discloses a

system and method 'for lexicon, segmentation and language model joint optimization' (col. 2,

lines 43-56), and teaches that 'a language model can take any sequence of items (**words,**

charters, letters, etc.) and estimate the probability of the sequence' (col. 1, line 35-41) and

providing 'a dynamic segmentation function 216 to segment items (**characters** or letters, for

example) into strings (e.g., words)', which suggests that the system/method has capability to

perform a character-based segmentation. Therefore, it would have been obvious to one of

ordinary skill in the art at the time the invention was made to recognize that the most popular

eastern languages, such as Chinese or Japanese, are character-based languages and have no blank

or space served as word boundaries in the written form, so that the combined system/method

from Wang in view of Razin can perform a segmentation for those character-based languages, as

Wang suggested, for the purpose (motivation) of improving language model performance and/or

providing capability of segmenting items (including characters) into words for a textural corpus

(Wang: col. 2, line 50-51 and col. 1, lines 52-59). This means that Wang in view of Razin,

alone, can also provides sufficient basis for the rejection, based on broadest reasonable

interpretation of the claim.

As per **claim 2** (depending on claim 1), Wang in view of Razin and Yang further

discloses "using punctuations, Arabic digits and alphabetic strings, or new words patterns to split

the cleaned corpus", (Razin, column 21, lines 10, 'punctuation'; column 4, lines 26, 'the usage of stop list');

As per **claim 3** (depending on claim 1), Wang in view of Razin and Yang further discloses "using common vocabulary to segment the cleaned corpus", (Razin: column 5, lines 36-45, 'the dictionary of standard phrases (common vocabulary)').

As per **claim 6** (depending on claim 1), Wang in view of Razin and Yang further discloses:

"filtering out functional words" (Razin: column 4, lines 35-38, 'stop list', 'semantically insignificant words (e.g., "and then about the") (interpreted as functional words)', which suggests that these words can be filtered out);

"filtering out those sub strings which almost always appear along with a longer sub strings" (Razin: column 9, lines 52, 'eliminates from the phrase list otherwise-significant phrases that are nested within other significant phrases... removes from the final phrase list minimal content words dangling at the beginning or end of preliminary user-specific phrases', which reads on the claim); and

"filtering out those sub strings for which the occurrence is less than a predetermined threshold", (Razin: column 2, lines 10-13, 'each node of tree is associated with a record of the number of occurrence of the word sequence at that node, where the number of occurrence exceeds the required threshold', which reads on the claimed limitation).

As per **claim 7** (depending on claim 1), Wang in view of Razin and Yang further discloses "using pre-recognized functional words as segment boundary patterns", (Razin: column

4, lines 35-38, 'stop list', 'semantically insignificant words (e.g., "and then about the")

(interpreted as functional words)').

As per **claim 8** (depending on claim 3), the rejection is based on the same reason

described for claim 7 because the claim recites the same or similar limitation(s) as claim 7.

As per **claim 9** (depending on claim 3), the rejection is based on the same reason

described for claim 6 because the claim recites the same or similar limitation(s) as claim 6.

As per **claims 10-12 and 15-18**, they recite an automatic new word extraction system.

The rejection is based on the same reason described for claims 1-3 and 6-9, respectively, because

the claims recite the same or similar limitation(s) as claims 1-3 and 6-9, respectively.

As per **claim 19**, it recites a program storage device readable by machine. The rejection

is based on the same reason described for claim 1, because the claim recites the same or similar

limitations as claim 1.

10.   Claims 4-5 and 13-14 are rejected under 35 U.S.C. 103(a) as being unpatentable over

Wang in view of Razin and Yang as applied to claims 1 and 10, and further in view of Hui (IDS:

"Color Set Size Problem with Applications to String Matching," Proc. of 2nd Symposium on

Combinatorial Pattern Matching, 1992, pp. 230-243).

As per **claim 4** (depending on claim 1), even Wang in view of Razin and Yang further

discloses using suffix tree (i.e. atomic suffix tree—AST) (Wang: column 1, line 42; Razin:

column, 2, line 3), Wang in view of Razin does not expressly disclose "using a GAST".

However, the feature is well known in the art as evidenced by Hui who teaches 'the concept of

suffix tree can be extended' and 'this extension is called the Generalized suffix tree (GST)(

corresponding to GAST)' (Hui, page 237, first paragraph). Therefore, it would have been

obvious to one of ordinary skill in the art at the time the invention was made to modify Wang in

view of Razin and Yang by specifically providing using extended suffix tree (GST or GAST), for

the purpose of storing more than one input strings (Hui: page 237, first paragraph).

As per **claim 5** (depending on claim 4), Wang in view of Razin, Yang and Hui further

discloses the tree "implemented by limiting length of sub strings", (Razin: column 14, lines 34-

35, 'length less than or equal to Smax').

As per **claim 13** (depending on claim 10), the rejection is based on the same reason

described for claim 4 because the claim recites the same or similar limitation(s) as claim 4.

As per **claim 14** (depending on claim 10), the rejection is based on the same reason

described for claim 5 because the claim recites the same or similar limitation(s) as claim 5.

Conclusion

11.   Please address mail to be delivered by the United States Postal Service (USPS) as
follows:

> Mail Stop ____
> Commissioner for Patents
> P.O. Box 1450
> Alexandria, VA 22313-1450

or faxed to: 571-273-8300, (for formal communications intended for entry)

Or: 571-273-8300, (for informal or draft communications, and please label
"PROPOSED" or "DRAFT")

If no Mail Stop is indicated below, the line beginning Mail Stop should be omitted from
the address.

Effective January 14, 2005, except correspondence for Maintenance Fee payments,
Deposit Account Replenishments (see 1.25(c)(4)), and Licensing and Review (see 37 CFR 5.1(c)
and 5.2(c)), please address correspondence to be delivered by other delivery services (Federal
Express (Fed Ex), UPS, DHL, Laser, Action, Purolater, etc.) as follows:

> U.S. Patent and Trademark Office
> Customer Window, Mail Stop _____

               Randolph Building
               Alexandria , VA 22314

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Qi Han whose telephone numbers is (571) 272-7604. The examiner can normally be reached on Monday through Thursday from 9:00 a.m. to 7:00 p.m. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil, can be reached on (571) 272-7602.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Inquiries regarding the status of submissions relating to an application or questions on the Private PAIR system should be directed to the Electronic Business Center (EBC) at 866-217-9197 (toll-free) or 703-305-3028 between the hours of 6 a.m. and midnight Monday through Friday EST, or by e-mail at: ebc@uspto.gov. For general information about the PAIR system, see http://pair-direct.uspto.gov.

QH/qh
August 9, 2006

RICHEMOND DORVIL
SUPERVISORY PATENT EXAMINER